Asymptotic normality of the size of the giant component in a random hypergraph

Béla Bollobás*† Oliver Riordan‡

December 15, 2011

Abstract

Recently, we adapted random walk arguments based on work of Nachmias and Peres, Martin-Löf, Karp and Aldous to give a simple proof of the asymptotic normality of the size of the giant component in the random graph $G(n,p)$ above the phase transition. Here we show that the same method applies to the analogous model of random k-uniform hypergraphs, establishing asymptotic normality throughout the (sparse) supercritical regime. Previously, asymptotic normality was known only towards the two ends of this regime.

1 Introduction and results

Let $H_k(n,p)$ denote the random k-uniform hypergraph with vertex set $[n] = \{1,2,\dots,n\}$ in which each of the $\binom{n}{k}$ possible edges is present independently with probability p. Thus $H_2(n,p)$ is the classical random graph $G(n,p)$. Our aim here is to study the component structure of $H_k(n,p)$, in particular the distribution of the size of the largest component above the 'phase transition'.

Before turning to the details, let us note that the notion of a 'component' in a k-uniform hypergraph H_k can be interpreted in a number of ways. For any $1 \leqslant r \leqslant k-1$, one could consider two edges to be 'connected' if they share at least r vertices (or perhaps exactly r vertices), and use this notion to define the components of H_k. Moreover, the size of a component could then be measured in a number of ways – either by the number of vertices that it contains, or (probably more naturally) by the number of r-sets of vertices, or the number of edges. In the rest of the paper we consider $r = 1$. It seems that

*Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK and Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA. E-mail: b.bollobas@dpmms.cam.ac.uk.

†Research supported in part by NSF grants DMS-0906634, CNS-0721983 and CCF-0728928, ARO grant W911NF-06-1-0076, and TAMOP-4.2.2/08/1/2008-0008 program of the Hungarian Development Agency

‡Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK. E-mail: riordan@maths.ox.ac.uk.

other values of r have received very little attention, although the corresponding notions of 'cycle' (with $r = 1$ being 'loose' and $r = k - 1$ being 'tight') have been studied extensively. Note that for hypergraphs derived from k-cliques in random graphs, questions about components defined using $r = k - 1$ were raised by Derényi, Palla and Vicsek [7]; such components were studied for all $1 \leqslant r \leqslant k-1$ in [3]. We shall say nothing further about the case $r \geqslant 2$, except to note that (greatly simplified versions of) the branching process arguments in [3] presumably show that the threshold for the emergence of a giant component is at $p \sim n^{r-k}(k-r)!/(\binom{k}{r} - 1)$.

For the rest of the paper we take $r = 1$, i.e., we say that two vertices are *connected* in a k-uniform hypergraph H if they are connected in the graph obtained by replacing each edge by a copy of K_k, and take the *components* of H to be the maximal sub-hypergraphs in which all vertices are connected. For reasons that will become clear below we write $p = p(n)$ as $\lambda(k-2)!n^{-k+1}$, where $\lambda = \lambda(n)$; when $k = 2$ this reduces to $p = \lambda/n$. Our main aim is to study the number L_1 of vertices in the largest component of $H_k(n, p)$ in the supercritical regime, i.e., when $(\lambda - 1)n^{1/3} \to \infty$. We also prove a result for the critical regime, where $(\lambda - 1)n^{1/3}$ is bounded.

For $k = 2$, very detailed results of this type are known. Pittel and Wormald [13] and Luczak and Łuczak [10] showed, in each case as part of a much stronger and/or more general result, that throughout the supercritical regime, L_1 is asymptotically normal: centralized and scaled appropriately, it converges in distribution to a standard normal distribution. The special case where $\lambda > 1$ is constant was proved earlier by Stepanov [14]. For hypergraphs, where $k \geqslant 3$ is fixed, much less is known: Karoński and Łuczak [8] proved strong results (a local limit theorem) in the barely supercritical phase, when $(\lambda - 1)^3 n$ tends to infinity but more slowly than $\log n / \log\log n$. At the other end of the range, Behrisch, Coja-Oghlan and Kang [2] proved a local limit theorem when $\lambda > 1$ is fixed. Here we shall prove asymptotic normality throughout the supercritical regime, for all $k \geqslant 3$ fixed. Note that our main result, while less precise than those of [8, 2], has a much greater range of applicability. The proof is a (to us surprisingly) simple adaptation of the argument we gave for the case $k = 2$ in [4], itself based on exploration and martingale arguments using ideas of Nachmias and Peres [12], Martin-Löf [11], Karp [9] and Aldous [1].

Given $\lambda > 1$, let λ_* be the 'dual branching process parameter', defined by $\lambda_* < 1$ and

$$\lambda_* e^{-\lambda_*} = \lambda e^{-\lambda}.$$

Writing $\mathfrak{X}_\mu$ for the Galton–Watson branching process in which the offspring distribution is Poisson with mean μ, it is well known that conditioning $\mathfrak{X}_\lambda$ on extinction gives $\mathfrak{X}_{\lambda_*}$. Let $\rho_\lambda = \rho_{2,\lambda}$ denote the survival probability of $\mathfrak{X}_\lambda$, so $\rho_\lambda > 0$ may be defined by

$$1 - \rho_\lambda = e^{-\lambda\rho_\lambda}, \tag{1}$$

and satisfies $\lambda_* = \lambda\rho_\lambda$. Finally, for $k \geqslant 3$ define $\rho_{k,\lambda}$ by

$$1 - \rho_{k,\lambda} = (1 - \rho_\lambda)^{1/(k-1)}; \tag{2}$$

it is easy to see that $\rho_{k,\lambda}$ is the survival probability of a certain branching process naturally associated to $H_k(n,p)$, where $p = \lambda(k-2)!n^{-k+1}$.

As usual, we say that a sequence (E_n) of events holds *with high probability* or *whp* if $\mathbb{P}(E_n) \to 0$ as $n \to \infty$. If (X_n) is a sequence of random variables and $f(n)$ is a deterministic function, then $X_n = o_{\mathrm{p}}(f(n))$ means that $X_n/f(n)$ converges to 0 in probability, i.e., that for any constant $\varepsilon > 0$, $|X_n| \leqslant \varepsilon f(n)$ holds whp. Later, we shall also use $X_n = O_{\mathrm{p}}(f(n))$ to mean that $X_n/f(n)$ is bounded in probability, i.e., for any $\varepsilon > 0$ there is a C such that for all (large enough) n we have $\mathbb{P}(|X_n| \geqslant Cf(n)) \leqslant \varepsilon$.

Writing $L_1(H)$ for the maximum number of vertices in any component of a hypergraph H, Coja-Oghlan, Moore and Sanwalani [6] showed that if $k \geqslant 3$ and $\lambda > 1$ are fixed and $p = p(n) = \lambda(k-2)!n^{-k+1}$, then $L_1(H_k(n,p)) = \rho_{k,\lambda}n + o_{\mathrm{p}}(n)$. (This result was certainly known as 'folklore' before this.) Our main result concerns the limiting distribution of the $o_{\mathrm{p}}(n)$ term, and applies throughout the supercritical regime.

Given $k \geqslant 2$ and $\lambda > 1$, let

$$\sigma_{k,\lambda}^2 = \frac{\lambda(1-\rho)^2 - \lambda_*(1-\rho) + \rho(1-\rho)}{(1-\lambda_*)^2} n, \tag{3}$$

where $\rho = \rho_{k,\lambda}$. It is well known that when $\lambda = 1+\varepsilon$ and $\varepsilon \to 0$, then $\rho_\lambda \sim 2\varepsilon$. From (2) it follows that

$$\rho_{k,\lambda} \sim \frac{2\varepsilon}{k-1},$$

and thus $\sigma_{k,\lambda}^2 \sim 2\varepsilon^{-1}n$. Expanding λ_* and thus ρ_λ and hence $\rho_{k,\lambda}$ further as series in $\varepsilon = 1-\lambda$, it is easy to check that in fact

$$\sigma_{k,\lambda}^2 = \left(2\varepsilon^{-1} + \frac{2(k-4)}{k-1} + O(\varepsilon)\right) n.$$

Thus, although the leading term does not depend on k, the next term does.

Theorem 1. *Let $k \geqslant 3$ be fixed, and let*

$$p = p(n) = \lambda(k-2)!n^{-k+1},$$

where $\lambda = \lambda(n)$ is bounded and $(\lambda-1)^3 n \to \infty$. Then

$$\frac{L_1(H_k(n,p)) - \rho_{k,\lambda}n}{\sigma_{k,\lambda}} \xrightarrow{\mathrm{d}} N(0,1),$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution, $N(0,1)$ is a standard normal random variable, and $\rho_{k,\lambda}$ and $\sigma_{k,\lambda}$ are defined in (2) *and* (3).

As a by-product of our proof, we obtain an analogue of the result of Aldous [1] giving the limiting distribution of the rescaled large component sizes inside the scaling window of the phase transition. Define a stochastic process $W^\alpha(s)$ (a random function on $[0,\infty)$) by

$$W^\alpha(s) = W(s) + \alpha s - s^2/2,$$

where $W(s)$ is a standard Brownian motion. As in [1], define an *excursion* of this process to be a maximal interval on which W^α exceeds its previous minimum value, and let $(|\gamma_i|)_{i\geqslant 1}$ denote the lengths of the excursions sorted into decreasing order. (Aldous shows that this makes sense with probability 1.)

Theorem 2. *Let $k \geqslant 3$ be fixed, and let $p = p(n) = \lambda(k-2)!n^{-k+1}$ where $\lambda = \lambda(n)$ satisfies*

$$(\lambda - 1)^3 n \to (k-1)^2\alpha^3$$

for some $\alpha \in \mathbb{R}$. Then, for any fixed r, writing L_r for the number of vertices in the rth largest component of $H(n,p)$, the sequence $((k-1)^{1/3}n^{-2/3}L_i)_{i=1}^r$ converges in distribution to $(|\gamma_i|)_{i=1}^r$ where $|\gamma_i|$ is defined as above.

The slightly strange scaling is chosen to match the graph case: the conclusion is that under these assumptions, up to a $(k-1)^{1/3}$ scaling factor, the large component sizes have the same limiting distribution as in the random graph $G(n,(1+\alpha n^{-1/3})/n)$.

2 Proofs

In this section we shall prove Theorems 1 and 2. The arguments, which closely follow those in [4], require a little preparation.

Let $H = H_k(n,p)$ be the random k-uniform hypergraph defined in the introduction. Our proofs are based on an algorithm for 'exploring' the components of H in n steps. For $0 \leqslant t \leqslant n$, 'time t' refers to the situation after t steps, so step t goes from time $t-1$ to time t. In step t we shall 'explore' a vertex v_t, meaning that we reveal all edges incident with v_t but not with any previously explored vertices. Noting that one vertex is explored in each step, this means that, however v_t is chosen, each of the $\binom{n-t}{k-1}$ possible edges containing v_t and not containing any previously explored vertices will be present with probability p, independently of the others and of the history.

More precisely, as in [4], at time t every vertex is either 'explored', 'active' or 'unseen'. We write A_t and U_t for the numbers of active and unseen vertices; exactly t vertices will be explored by time t, so $A_t + U_t = n - t$. At time $t = 0$, we have $A_0 = 0$ and $U_0 = n$. Fix an order on the vertices. In step $1 \leqslant t \leqslant n$ we choose v_t to be the first active vertex (at time $t-1$), if there are any; otherwise v_t is the first unseen vertex. In the latter case we say that we 'start a new component' in step t. In step t we reveal all edges containing v_t and not containing any explored vertex. Let η_t be the number of unseen vertices other than v_t in such edges. These η_t vertices are now labelled active (at time t), and v_t is labelled as explored. It is easy to check that the process reveals the components of H one-by-one, starting a new component in step t whenever $A_{t-1} = 0$. Thus, if $0 = t_0 < t_1 < t_2 < \cdots < t_k = n$ enumerates $\{t : A_t = 0\}$, then the sequence $(t_i - t_{i-1})_{i=1}^k$ lists exactly the numbers of vertices in the components of H, in some order. In particular,

$$L_1 = \max\{t_i - t_{i-1} : 1 \leqslant i \leqslant k\}. \tag{4}$$

We shall study the random walk (X_t) defined by $X_t = A_t - C_t$, where C_t is the number of new components started within the first t steps. As in [4], we have $t_i = \inf\{t : X_t = -i\}$. In step t, exactly η_t vertices change state from unseen to active. Furthermore, one vertex v_t changes its state to explored: if a new component is started in step t, then v_t was previously unseen, otherwise it was active. Hence, if we start a new component, then $A_t = A_{t-1} + \eta_t$, otherwise $A_t = A_{t-1} + \eta_t - 1$. Since $A_0 = C_0 = 0$, it follows that

$$X_t = A_t - C_t = \sum_{i=1}^{t} (\eta_i - 1).$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting our random hypergraph $H_k(n,p)$. (For example, take Ω to be the set of all $2^{\binom{n}{k}}$ k-uniform hypergraphs on $[n]$, $\mathcal{F}$ to be the power-set of Ω, and $\mathbb{P}$ to be the appropriate probability measure.) Let $\mathcal{F}_t \subseteq \mathcal{F}$ denote the sub-sigma-field generated by all information revealed by time t. Following the strategy of [4], the key task is to understand the distribution of η_{t+1} given $\mathcal{F}_t$. Crucially, it is only the expectation that we need to bound precisely; our bound on the variance can be much cruder.

Given $\mathcal{F}_t$, we know which vertex v_{t+1} we are about to explore in step $t+1$. At time t there are $U'_t = U_t - 1_{\{A_t=0\}} = U_t - (C_{t+1} - C_t)$ unseen vertices u other than v_{t+1}. (The vertex v_{t+1} is active if $A_t > 0$; otherwise it is unseen.) For each such unseen vertex u there are exactly

$$c_{t+1} = \binom{A_t + U_t - 2}{k-2} = \binom{n-t-2}{k-2} \tag{5}$$

potential edges containing v_{t+1} and u but not containing any of the t vertices previously explored. Since each such edge is present with probability p, the probability that u becomes active during step $t+1$ is

$$\pi_1 = 1 - (1-p)^{c_{t+1}} = pc_{t+1} + O(p^2 c_{t+1}^2) = \lambda (n-t)^{k-2} n^{-k+1} + O(1/n^2).$$

Of course, these events are not independent for different u, but this does not matter for the expectation. In particular,

$$\mathbb{E}(\eta_{t+1} \mid \mathcal{F}_t) = U'_t \pi_1 = U'_t p c_{t+1} + O(1/n). \tag{6}$$

Fortunately for the subsequent analysis, this expression depends on U'_t in a linear way.

We next estimate the conditional variance of η_{t+1} given $\mathcal{F}_t$; here we do not need to be so accurate. Let u_1 and u_2 be distinct unseen vertices (other than v_{t+1} if that happens to be unseen). The probability that u_1 and u_2 both become active is $\pi_2 + \pi_3$, where π_2 is the probability that we find an edge containing v_{t+1}, u_1 and u_2, and π_3 is the probability that this does not happen, but we find disjoint edges activating u_1 and u_2. Supposing that $n - t \to \infty$, then

$$\pi_2 = 1 - (1-p)^{\binom{n-t-3}{k-3}} \sim p \binom{n-t-3}{k-3} \sim \lambda (k-2)(n-t)^{k-3} n^{-k+1}.$$

Also, it is easy to check that $\pi_3 \sim \pi_1^2$. Hence,

$$\begin{aligned}\mathrm{Var}(\eta_{t+1} \mid \mathcal{F}_t) &= U_t'(U_t'-1)(\pi_2+\pi_3) + U_t'\pi_1 - (U_t'\pi_1)^2 \\ &\sim (U_t')^2\pi_2 + U_t'\pi_1 \\ &\sim \lambda(k-2)(1-t/n)^{k-3}\frac{U_t^2}{n^2} + \lambda(1-t/n)^{k-2}\frac{U_t}{n}. \qquad (7)\end{aligned}$$

In particular, the maximum possible value satisfies

$$\max_t \sup_\Omega \mathrm{Var}(\eta_{t+1} \mid \mathcal{F}_t) \leqslant \lambda(k-1) = O(1). \qquad (8)$$

Let $D_t = \mathbb{E}(\eta_t - 1 \mid \mathcal{F}_{t-1})$. Recalling that $U_t = n-t-A_t = n-t-(X_t+C_t)$, and noting that $U_t' = U_t - (C_{t+1} - C_t) = n - t - X_t - C_{t+1}$, from (6) we see that

$$D_{t+1} = pU_t'c_{t+1} - 1 + O(1/n) = \alpha_{t+1}(n-t-X_t-C_{t+1}) - 1 + O(1/n), \quad (9)$$

where

$$\alpha_t = pc_t = p\binom{n-t-1}{k-2}.$$

Set $\Delta_{t+1} = X_{t+1} - X_t - D_{t+1}$, so Δ_{t+1} is $\mathcal{F}_{t+1}$-measurable and $\mathbb{E}(\Delta_{t+1} \mid \mathcal{F}_t) = 0$, by the definition of D_{t+1}. Note that

$$\begin{aligned}X_{t+1} &= X_t + D_{t+1} + \Delta_{t+1} \\ &= (1-\alpha_{t+1})X_t + \alpha_{t+1}(n-t) - 1 + \Delta_{t+1} - \alpha_{t+1}C_{t+1} + O(1/n). \qquad (10)\end{aligned}$$

We shall approximate (X_t) by the sum of a deterministic sequence and a martingale. To this end, define a deterministic sequence (x_t) by $x_0 = 0$ and

$$x_{t+1} = (1-\alpha_{t+1})x_t + \alpha_{t+1}(n-t) - 1. \qquad (11)$$

Subtracting (11) from (10) we see that

$$X_{t+1} - x_{t+1} = (1-\alpha_{t+1})(X_t - x_t) + \Delta_{t+1} - \alpha_{t+1}C_{t+1} + E_{t+1}, \qquad (12)$$

where E_{t+1} is an 'error term' with $E_{t+1} = O(1/n)$. Defining

$$\beta_t = \prod_{i=1}^{t}(1-\alpha_i),$$

the recurrence relation (12) may be easily solved to give

$$X_t - x_t = \sum_{i=1}^{t}\frac{\beta_t}{\beta_i}(\Delta_i - \alpha_i C_i + E_i). \qquad (13)$$

Note that $0 < \alpha_i < 1$ for each i, so the sequence β_t is decreasing.

Motivated by this formula, let

$$S_t = \sum_{i=1}^{t} \beta_i^{-1} \Delta_i, \tag{14}$$

so (S_t) is a martingale with respect to $(\mathcal{F}_t)$, and set

$$\widetilde{X}_t = x_t + \beta_t S_t :$$

this is our desired approximation for (X_t).

Lemma 3. *For any $p = p(n) = O(n^{-k+1})$ we have*

$$|X_t - \widetilde{X}_t| = O(tC_t/n),$$

uniformly in $1 \leqslant t \leqslant n$.

Proof. From (13) and the definition of $\widetilde{X}_t$ we have

$$X_t - \widetilde{X}_t = \sum_{i=1}^{t} \frac{\beta_t}{\beta_i}(E_i - \alpha_i C_i).$$

The result follows from the fact that $\beta_t/\beta_i = \prod_{j=i+1}^{t}(1-\alpha_j)$ is between 0 and 1, the bounds $E_i, \alpha_i = O(1/n)$, and the fact that $C_i \leqslant C_t$ for $i \leqslant t$. □

We next analyze the deterministic trajectory (x_t). Setting $x_t = n - t - y_t$, the relation (11) can be rearranged to give $y_{t+1} = (1-\alpha_{t+1})y_t$. Since $y_0 = n - x_0 = n$, we see that $y_t = n\beta_t$, so

$$x_t = n - t - n\beta_t. \tag{15}$$

Recall that the α_i are $O(1/n)$. Since there are at most n terms in the sum, it follows that

$$\log \beta_t = \sum_{i=1}^{t} \log(1-\alpha_i) = -\sum_{i=1}^{t} \alpha_i + O(1/n).$$

From the definition of $\alpha_i = pc_i$ we have

$$\sum_{i=1}^{t} \alpha_i = p \sum_{i=1}^{t} \binom{n-i-1}{k-2} = p\left(\binom{n-1}{k-1} - \binom{n-t-1}{k-1}\right).$$

It follows that

$$\begin{aligned} \log \beta_t &= -\frac{pn^{k-1}}{(k-1)!}\left(1 - (1-t/n)^{k-1}\right) + O(1/n) \\ &= -\frac{\lambda}{k-1}\left(1 - (1-t/n)^{k-1}\right) + O(1/n). \end{aligned} \tag{16}$$

Define the function $g = g_{k,\lambda}$ by

$$g_{k,\lambda}(\tau) = 1 - \tau - \exp\left(-\frac{\lambda}{k-1}\Big(1-(1-\tau)^{k-1}\Big)\right) \tag{17}$$

and (for compatibility with the notation in [4]) set

$$f(t) = f_{n,k,\lambda}(t) = ng(t/n).$$

Then (15) and (16) imply that that $x_t = f(t) + O(1)$, uniformly in $0 \leqslant t \leqslant n$. In other words, the function f or g represents a (rescaled in the case of g) idealized form of the deterministic approximation (x_t) to (X_t).

From (1) and (2) it is easy to check that $\rho = \rho_{k,\lambda}$ satisfies $g(\rho) = 0$. Note that

$$g'(\tau) = -1 + \lambda(1-\tau)^{k-2}\exp\left(-\frac{\lambda}{k-1}\Big(1-(1-\tau)^{k-1}\Big)\right),$$

so $g'(0) = \lambda - 1$. Also, recalling that $(1-\rho)^{k-1} = 1 - \rho_\lambda = \lambda_*/\lambda$,

$$g'(\rho) = -1 + \lambda(1-\rho)^{k-1} = -(1-\lambda_*). \tag{18}$$

Furthermore,

$$g''(\tau) = \Big(-\lambda(k-2)(1-\tau)^{k-3} - \big(\lambda(1-\tau)^{k-2}\big)^2\Big) \exp\left(-\frac{\lambda}{k-1}\Big(1-(1-\tau)^{k-1}\Big)\right),$$

so $g'' \leqslant 0$. Hence g is concave, so f is concave. Also, $\sup\{|g''(\tau)| : 0 \leqslant \tau \leqslant 1\} = O(1)$, so $f''(t) = O(1/n)$, uniformly in $0 \leqslant t \leqslant n$.

Note that

$$g''(0) = -\big(\lambda(k-2) + \lambda^2\big) = -(k-1) + O(\varepsilon),$$

where $\varepsilon = \lambda - 1$. Since (as is easily checked) g''' is uniformly bounded, it follows that for $\tau = O(\varepsilon)$ we have

$$g(\tau) = g(0) + \tau g'(0) + \tau^2 g''(0)/2 + O(\tau^3) = \varepsilon\tau - (k-1)\tau^2/2 + O(\varepsilon^3). \tag{19}$$

With this preparation behind us, the proof of Theorem 1 will follow that in [4] very closely, so we give only an outline. We use the same notation as in [4] whenever possible.

Proof of Theorem 1. Firstly, we have $\log \beta_t = O(1)$ uniformly in $0 \leqslant t \leqslant n$. This and (8) imply that the increments $\beta_i^{-1}\Delta_i$ of the martingale (S_t) have variance $O(1)$, so $\mathrm{Var}(S_t) = O(t)$ for any deterministic $t = t(n)$. Hence, by Doob's maximal inequality, $\max_{i \leqslant t} |S_t| = O_{\mathrm{p}}(\sqrt{t})$. In particular,

$$\widetilde{X}_t = x_t + \beta_t S_t = f(t) + \beta_t S_t + O(1) = f(t) + O_{\mathrm{p}}(\sqrt{t}). \tag{20}$$

Let $\sigma_0 = \sqrt{\varepsilon n}$, let $\omega = \omega(n)$ tend to infinity slowly (with $\omega^6 = o(\varepsilon^3 n)$), and let $t_0 = \omega\sigma_0/\varepsilon$. Write $Z = -\inf\{X_t : t \leqslant t_0\}$ for the number of components completely explored by time t_0, and T_0 for the time at which we finish exploring the last such component. Since $f'(0) = g'(0) = \varepsilon$, and we have the bound $X_t - \widetilde{X}_t = O(tC_t/n)$ from Lemma 3, the proof of [4, Lemma 6] goes through *mutatis mutandis* to show that $Z \leqslant \sigma_0/\omega$ and $T_0 \leqslant \sigma_0/(\varepsilon\omega)$. Note for later that the latter bound gives $T_0 = o_{\mathrm{p}}(\sqrt{n/\varepsilon})$.

Writing $t_1 = \rho_{k,\lambda} n$ (ignoring the irrelevant rounding to integers), let T_1 be the first time at which we finish exploring a component after time t_0. Arguing exactly as in [4] we see that $t_1 - t_0 \leqslant T_1 \leqslant t_1 + t_0$ holds whp, and indeed that

$$T_1 = t_1 + \widetilde{X}_{t_1}/(1-\lambda_*) + o_{\mathrm{p}}(\sigma_0/\varepsilon). \tag{21}$$

(Relation (21) takes the place of equation (21) in [4].)

We claim that for each $t \leqslant t_1$ we have $U_t = u_t + o_{\mathrm{p}}(n)$, where now

$$u_t = n \exp\left(-\frac{\lambda}{k-1}\Big(1 - (1-t/n)^{k-1}\Big)\right). \tag{22}$$

Noting that $u_t = n - t - f(t)$, and recalling that $U_t = n - t - A_t = n - t - (X_t + C_t)$, this can be deduced from the crude bound (20) on $\widetilde{X}_t$ and Lemma 3 by arguing as in [4]. Note that, as pointed out to us (in the graph case) by Lutz Warnke, the approximate form of this formula is easy to guess: ignoring vertices selected to start new components, a given vertex u is unseen at time t if and only if none of the potential edges containing u tested during the first t steps was found to be present. There are $\binom{n-1}{k-1} - \binom{n-t-1}{k-1}$ such edges: those containing u and at least one of the t explored vertices.

From (7) and (22), the sum of the conditional variances of the first t_1 increments of (S_t) is concentrated around

$$\sum_{i=1}^{t_1} \beta_i^{-2} \left(\lambda(k-2)(1-i/n)^{k-3}(u_i/n)^2 + \lambda(1-i/n)^{k-2} u_i/n\right). \tag{23}$$

Although the distribution of the increments is not quite as nice as in the graph case, it is easy to see that the conditional distribution of η_t given $\mathcal{F}_{t-1}$ is dominated by $k-1$ times a binomial random variable with mean $O(1)$. (The binomial random variable is the number of edges found; each contributes a number of new unseen vertices between 0 and $k-1$.) Thus any fixed moment of η_t is bounded by a constant, and this transfers to $D_t = \eta_t - \mathbb{E}(\eta_t \mid \mathcal{F}_{t-1})$ and hence to $\beta^{-t} D_t$. This condition (for the fourth moment) and concentration of the sum of the conditional variances is more than enough for a martingale central limit theorem such as Brown [5, Theorem 2], and it follows that S_{t_1}, and hence $\widetilde{X}_{t_1}$ and thus T_1, is asymptotically normally distributed.

For the variance, the sum in (23) is well approximated by an integral, and after a slightly unpleasant calculation one sees that

$$\mathrm{Var}(\widetilde{X}_{t_1}) = \mathrm{Var}(\beta_t S_t) \sim \big(\lambda(1-\rho)^2 - \lambda_*(1-\rho) + \rho(1-\rho)\big)n = (1-\lambda_*)^2\sigma^2,$$

where $\rho = \rho_{k,\lambda}$ and $\sigma = \sigma_{k,\lambda}$ are defined in (2) and (3). Recalling (21), and noting that $T_0 = o_{\mathrm{p}}(\sqrt{n/\varepsilon}) = o_{\mathrm{p}}(\sigma)$, it follows that T_1 and thus $T_1 - T_0$ is asymptotically normal with mean ρn and variance σ^2, so there is a component whose size has the required distribution.

Finally, it is not hard to check that there is whp no larger component, using the fact that what remains to be explored after time T_1 is a subcritical random hypergraph. Specifically, one can either apply a martingale argument as in Nachmias and Peres [12], or simply apply the subcritical case of the results proved by Karoński and Łuczak [8]. □

Proof of Theorem 2. Suppose that $p = \lambda(k-2)!n^{-k+1}$ with $\lambda = 1+\varepsilon$, where $\varepsilon = \varepsilon(n)$ satisfies $\varepsilon n^{1/3} \to (k-1)^{2/3}\alpha$ as $n \to \infty$, for some $\alpha \in \mathbb{R}$ constant. As before we consider the random walk (X_t), but now only for $t \leqslant An^{2/3}$ for a large constant A. Aldous [1] shows in the graph case that, appropriately rescaled, the process (X_t) converges to a deterministic quadratic plus a standard Brownian motion. We show the same here, with slightly different rescaling.

The argument giving $x_t = ng(t/n)+O(1)$ above assumed only that $\lambda = O(1)$, which applies here. Writing $t = s(k-1)^{-1/3}n^{2/3}$, using (19) we see that for $t \leqslant An^{2/3}$ we have

$$\begin{aligned} x_t/((k-1)n)^{1/3} &= (k-1)^{-1/3}n^{2/3}g\big(s(k-1)^{-1/3}n^{-1/3}\big) + o(1) \\ &= n^{1/3}\varepsilon(k-1)^{-2/3}s - s^2/2 + o(1) = \alpha s - s^2/2 + o(1). \end{aligned}$$

In other words, the deterministic limiting trajectory is quadratic. Moreover, in this range (indeed, whenever $t = o(n)$) we have $\beta_t \sim 1$ (see (16)) and, from (7), the martingale differences appearing in (14) have (conditional) variance

$$\beta_i^{-2}\mathrm{Var}(\Delta_i \mid \mathcal{F}_{i-1}) \sim \mathrm{Var}(\Delta_i \mid \mathcal{F}_{i-1}) = \mathrm{Var}(\eta_i \mid \mathcal{F}_{i-1}) \sim k-1.$$

It follows using the bounds on $|X_t - x_t - S_t|$ established above that the rescaled process whose value at rescaled time s is $X_{s(k-1)^{-1/3}n^{2/3}}/((k-1)n)^{1/3}$ converges to the quadratic function above plus a standard Brownian motion, i.e., to $W^\alpha(s)$. The rest of the argument is exactly as in the original paper of Aldous [1], so we omit the details, noting only that it is the *time* rescaling factor that appears in the component sizes. □

For slightly more details of a related argument, see [15, Section 4].

Acknowledgement. This research was triggered by a question that Tomasz Łuczak asked us during the conference Random Structures and Algorithms 2011 in Atlanta. We are grateful to him for raising this question.

References

[1] D. Aldous, Brownian excursions, critical random graphs and the multiplicative coalescent, *Ann. Probab.* **25** (1997), 812–854.

[2] M. Behrisch, A. Coja-Oghlan and M. Kang, The order of the giant component of random hypergraphs, *Random Struct. Alg.* **36** (2010), 149–184.

[3] B. Bollobás and O. Riordan, Clique percolation, *Random Struct. Alg.* **35** (2009), 294–322.

[4] B. Bollobás and O. Riordan, Asymptotic normality of the size of the giant component via a random walk *J. Combinatorial Theory B* **102** (2012), 53–61.

[5] B.M. Brown, Martingale central limit theorems, *Ann. Math. Stat.* **42** (1971), 59–66.

[6] A. Coja-Oghlan, C. Moore and V. Sanwalani, Counting connected graphs and hypergraphs via the probabilistic method, *Random Struct. Alg.* **31** (2007), 288–329.

[7] I. Derényi, G. Palla and T. Vicsek, Clique percolation in random networks, *Physical Review Letters* **94** (2005), 160202 (4 pages).

[8] M. Karoński and T. Łuczak, The phase transition in a random hypergraph, *J. Comput. Appl. Math.* **142** (2002), 125–135.

[9] R.M. Karp, The transitive closure of a random digraph, *Random Structures Algorithms* **1** (1990), 73–93.

[10] M. Luczak and T. Łuczak, The phase transition in the cluster-scaled model of a random graph, *Random Structures Algorithms* **28** (2006), 215246.

[11] A. Martin-Löf, Symmetric sampling procedures, general epidemic processes and their threshold limit theorems, *J. Appl. Probab.* **23** (1986), 265–282.

[12] A. Nachmias and Y. Peres, Component sizes of the random graph outside the scaling window, *ALEA Lat. Am. J. Probab. Math. Stat.* **3** (2007), 133–142.

[13] B. Pittel and C. Wormald, Counting connected graphs inside-out, *J. Combinatorial Theory B* **93** (2005), 127–172.

[14] V.E. Stepanov, Phase transitions in random graphs. (Russian) *Teor. Verojatnost. i Primenen.* **15** (1970), 200–216. Translated in Theory Probab. Appl. 15 (1970), 55–67.

[15] O. Riordan, The phase transition in the configuration model, to appear in *Combin. Probab. Comput.*, arXiv preprint 1104.0613